Rule-based Retrieval of Human Motion Data Using Inductive Logic Programming

Tomohiko Mukai Ken-ichi Wakisaka Shigeru Kuriyama *
Toyohashi University of Technology

Abstract

This paper proposes a novel method for automatically classifying and retrieving motions based on spatio-temporal features of motion appearance. Our method first converts a motion data into a form of clausal language that represents geometrical relations between human body parts and their temporal relations. A classification rule is then learned from the minimal set of manually classified examples using inductive logic programming. We introduce a two-step search algorithm to retrieve motion segments from the motion database, which uses two types of classification rules that are discovered by different learning models. Our method allows robust and efficient retrieval from complex motion sequences with a small number of training data.

CR Categories: I.3.7 [Computer Graphics]: Animation—

Keywords: motion retrieval, geometric feature, inductive logic programming

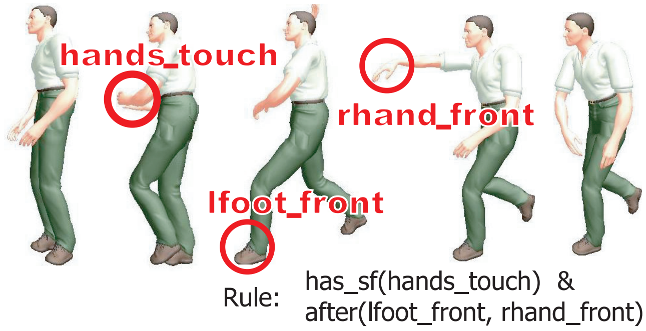

Figure 1: *The classification rule of throwing motion discovered by our method, represented in the clausal form. The motion segment is retrieved from a long motion sequence using the classification rule.*

1 Introduction

Automatic retrieval of human motion data from a large dataset has a potential benefit for creating character animation using a motion editing technique. Existing motion retrieval methods often use a short motion clip as the retrieval key. This technique defines the numerical distance measure to quantify the similarity between the retrieval key and data entities. However, such methods do not take semantic similarity into account in the search. For example, although both overarm and underarm throws can be classified as throwing motions, the overarm throw cannot be retrieved using an underarm throw as the retrieval key because their appearances are quite different.

In the fields of artificial intelligence and data mining, a general induction technique has been developed to discover an essential classification rule from particular examples. The induction method uses a clausal language to represent the example data, and discovers a fundamental classification rule from the examples using logical programming, called inductive logic programming (ILP) [Nienhuys-Cheng and de Wolf 1997]. Numerical methods, such as support vector machine and state-space models, implicitly learn the classifier from pre-classified training data, so it is difficult to modify the feature of classifier after the learning. In contrast, ILP learns the classification rule presented in the near natural language form without using any numerical objective function.

We propose a novel motion retrieval method using ILP. Our method first computes a set of spatio-temporal features of motion appearance in the form of a clausal language. An ILP framework then discovers an essential classification rule by analyzing an intrinsic difference among the minimal set of training motion clips. The classification rule is composed of a few logical expressions as shown in Figure 1, and they are easily translated into natural language that enables manual editing of the classification rule. The discovered rule is used to retrieve desirable segments from complex motion sequences. Consequently, our system provides more accurate and robust retrieval than existing numerical methods using a small training dataset.

*e-mail: {mukai@, k_wakisaka@val., kuriyama@}ics.tut.ac.jp

2 Inductive Learning of Classification Rules

Given training examples that are manually classified into several classes, one class is chosen as a positive class and others are used as a negative class. The inductive learning from the two classes discovers a classification rule in the same clausal form so that the resulting rule explains common features of the positive examples and no features of the negative examples.

The spatio-temporal features of training examples are converted into a clausal form. In order to reduce the computational cost by removing unessential oscillation of human motion, several key-poses are first extracted from a training motion clip by a greedy search. Next, the multilevel spatial features are computed at each key-pose and then represented in the clausal form like $has\_sf(M, f, l_f)$, which means a motion M has a spatial feature f of level l_f. The feature level is determined by quantizing the geometrical distance between body parts. For example, the feature level of *right_foot_up* is determined according to the height of the right foot from the ground. We currently use a five-level quantization for every multilevel feature. Our definition of spatial feature includes 31 geometrical features proposed in [Müller et al. 2005] and additional 4 customized features. We also define 5 temporal features that explain the duration of a spatial feature and a temporal relation between different spatial features.

The ILP framework automatically discovers the classification rule for each motion class that is represented in the clausal form. We use a public ILP system, called Progol [Muggleton 1995], which provides two types of learning model: **Exclusive learning model** uses both positive and negative examples to discover a rule that is obeyed by the positive examples and is violated by the negative examples, and **Positive learning model** discovers the rule using only positive examples based on Bayes theory using pseudo-negative examples that are generated by adding artificial noise to the positive examples. The existing method [Müller and Röder 2006] creates a motion template only from motions in the same class, which can be categorized as positive learning algorithms, whereas ILP can introduce negative examples into the learning to discover a rule having a higher classification capability.

3 Rule-based Motion Retrieval

Motion segments are retrieved by a subsequence search using a classification rule in a manner similar to template matching. The target sequence is first converted into a timeseries of spatial features. The subsequences are compared with each clause of the exclusive rule, and the matched segments are labeled as retrieval candidates. After the motion sequence is scanned by exclusive rules of all motion classes, the second search with a positive rule is processed by the same procedure on the subsequence that has not yet been annotated. Finally, successive segments having same label are joined to compose a retrieval result.

The rule-based retrieval can use a motion clip as the search query. Given a query motion clip, every classification rule is checked if it categorizes the query motion into one of the given classes, and the applied rules are then used for retrieving the similar motion segments. Even though the actual retrieval key is not the query motion itself, the classification rule is associated with the query. This approach enables the retrieval of a semantically similar motion with a large difference in appearance. This property can overcome the limitations in existing numerical techniques based on visual similarity. On the other hand, the existing methods use query motion clip only for comparing visual similarity.

4 Experimental Results

We experimentally retrieved motion segments from a long sequence of motions. The training dataset includes seven motion classes: *walk*(11), *run*(16), *jump*(6), *throw*(6), *lie*(6), *lift*(6), and *sit*(6), where the number in () denotes the number of motion samples. The classification rule for each motion class is discovered from a training dataset using exclusive and positive learning. We compare our method to a baseline method based on appearance feature matching [Müller and Röder 2006]. The baseline method creates a motion template as a retrieval key for each motion class by the weighted average of training motion clips. We create seven motion templates from the same training dataset.

We demonstrate the retrieval results for a complex motion sequence. The target sequence consists of training examples in which walking and sitting movements are alternately repeated 7 times. Figure 2 illustrates the search results by manual operation and the two methods. It shows that most is appropriately retrieved by our method the same as for manual operation. The false retrieval of *Lift* derived from a squatting movement commonly included in both sitting and lifting motion, and this defect is crucial for the baseline method.

Figure 3 shows the results of different motion sequences performed by another actor. The target sequence is well annotated by our method with less false retrieval. However, three segments of walking motion are passed over because they have different feature levels of knee bending angle and length of stride. Thus, the error of the logic-based analysis might be caused by the mis-discretization of the feature level, which should be improved in future work.

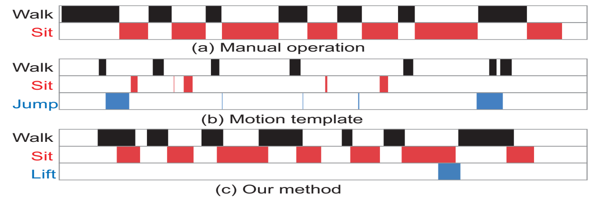

Figure 2: *Retrieval result of a walking and sitting motion sequence whose elemental motions are all used as training example.*

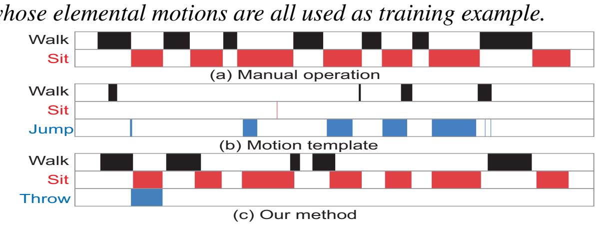

Figure 3: *Retrieval result of a walking and sitting motion sequence performed by another actor.*

5 Conclusions

This paper has proposed a motion retrieval method using inductive logic programming. The clausal formulation provides a meaningful representation of human motion and its classification rules. The exclusive and positive classification rules are efficiently learned within the ILP framework from a minimal set of training examples. By specifying the name of a motion class, motion segments are efficiently retrieved with a two-step search using the exclusive and positive classification rules, which enable robust retrieval from a complex motion sequence. Our system also retrieves the motions by using a short motion clip as the search query, which considers the semantic similarity between motions because the retrieval is actually done by using the classification rule extracted from the query motion.

The major limitation of our method is that the inductive learning requires a significant amount of computations. Several key-frames are therefore greedily sampled instead of using all frames in a motion sequence, but this simplification often loses important features of motions. The trade-off can be solved by introducing a more effective representation of appearance features.

References

MUGGLETON, S. 1995. Inverse entailment and progol. *New Generation Computing 13*, 245–286.

MUGGLETON, S. 1996. Learning from positive data. In *Sixth International Workshop on Inductive Logic Programming*, 358–376.

MÜLLER, M., AND RÖDER, T. 2006. Motion templates for automatic classification and retrieval of motion capture data. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation 2006*, 137–146.

MÜLLER, M., RÖDER, T., AND CLAUSEN, M. 2005. Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics 24*, 3, 677–685.

NIENHUYS-CHENG, S.-H., AND DE WOLF, R. 1997. *Foundation of Inductive Logic Programming*. Lecture Notes in Artificial Intelligence 1228. Springer.